

A Novel Reconfigurable Hardware Design for Speech Enhancement Based on Multi-Band Spectral Subtraction Involving Magnitude and Phase Components

Tanmay Biswas¹, Sudhindu Bikash Mandal¹, Debasree Saha², Amlan Chakrabarti³

IEEE Student Member¹, IEEE Member², IEEE Senior Member³

A.K.Choudhuri School Of Information Technology, University Of Calcutta^{1,2,3}

(tanmay123g, sudhindu.mandal)@gmail.com¹, debasri_cu@yahoo.co.in², acakcs@caluniv.ac.in³

August 26, 2015

Abstract

This paper proposes an efficient reconfigurable hardware design for speech enhancement based on multi band spectral subtraction algorithm and involving both magnitude and phase components. Our proposed design is novel as it estimates environmental noise from speech adaptively utilizing both magnitude and phase components of the speech spectrum. We performed multi-band spectral subtraction by dividing the noisy speech spectrum into different non-uniform frequency bands having varying signal to noise ratio (SNR) and subtracting the estimated noise from each of these frequency bands. This results to the elimination of noise from both high SNR and low SNR signal components for all the frequency bands. We have coined our proposed speech enhancement technique as Multi Band Magnitude Phase Spectral Subtraction (MBMPSS). The magnitude and phase operations are executed concurrently exploiting the parallel logic blocks of Field Programmable Gate Array (FPGA), thus increasing the throughput of the system to a great extent. We have implemented our design on Spartan6 Lx45 FPGA and presented the implementation result in terms of resource utilization and delay information for the different blocks of our design. To the best of our best knowledge, this is a new type of design for speech enhancement application and also a first of its kind implementation on reconfigurable hardware. We have used benchmark audio data for the evaluation of the proposed hardware and the experimental results show that our hardware shows a better SNR value compared to the existing state of the art research works.

Keyword's: Spectral Subtraction, Multi Band Spectral Subtraction, Digital Signal Processing (DSP), Field Programmable Gate Array (FPGA), System Generator.

1 Introduction

Speech enhancement aims to improve the quality of speech in a noisy environment. The spectral subtraction technique is a well known technique for speech noise elimination, which was originally introduced by S. Boll [1]. An upgraded version was introduced by Berouti et al. [2] for the musical noise reduction. The general principle behind the spectral subtraction is to estimate noise from the magnitude spectrum, which then gets subtracted from the original signal keeping the phase part of the spectrum unchanged. This general spectral subtraction technique results to three kinds of error [3] viz. error in noise estimation, error due to ignoring the speech-noise cross term in magnitude spectrum and error due to noisy phase spectrum with clear magnitude spectrum in signal reconstruction. The performance of speech enhancement is put down due to these errors. These errors has been reported in [4] [5] [6] for speech enhancement and speech recognition methods. When SNR of the signal is high, the noisy phase is close to the clean phase and the above methods work properly. But, when SNR drops then the cross term errors are produced, and the phase of the noisy signal plays the more seeming role in the clean magnitude signals and affects the reconstruction process. Recently, real and imaginary modulation spectral subtraction for speech enhancement was introduced by Yi Zhang et al. [3], where the subtraction procedure performed on both real and imaginary parts of the spectrum. Also the real world noise affects signal in various time intervals, which is also called colored noise. The real world noise spectrum are not flat like white Gaussian noise. The multi band spectral subtraction method for speech enhancement was introduced by Kamath [7], where the spectrum was divided into several bands for efficient noise reduction. In [8], design of multi band spectral subtraction was proposed based

on the magnitude compensation and phase modification. In [9], we can find a phase based dual microphone algorithm for robust speech enhancement. Plenty of research work based on spectral subtraction algorithm can be found in [11] [12] [13] [14] [15] [16] [17]. Speech enhancement based on hardware software co-design using FPGA platform can be found in [10] [18] [19] [20].

In this paper, we have performed noise estimation from both magnitude and phase spectrum by dividing the whole noisy speech spectrum into different non-uniform linearly shaped frequency bands and then subtracted the estimated noise from each frequency bands with different SNR over subtraction factor values α . This correctly justifies that our proposed hardware design for speech enhancement performs well for both high SNR signal as well as low SNR signal with different frequency bands. The hardware execution can be carried out in two ways: (a) off the shelf Digital Signal Processors (DSPs) and (b) FPGAs. We have chosen FPGA as our target hardware as it gives the opportunity of parallel computing involving the configurable logic cells [21] and dedicated DSP blocks. This leads to faster execution of hardware tasks, satisfying our primary objective. We have used the Xilinx System Generator tool in the MATLAB/SIMULINK environment [22] to design and verify our hardware. Here, we convey the comparative experimental results of SNR performance of the proposed architecture against the Magnitude Spectral Subtraction (MSS) [1], Magnitude Phase Spectral Subtraction (MPSS) [3] and Multi Band Magnitude Spectral Subtraction (MBMSS) [7] for different noisy signals, which clearly infers that our design yields better performance. We also convey the resource utilization and delay information of the proposed architecture. The major contributions of this work can be summarized as follows:

1. Proposal of a new speech enhancement method, relatively more robust as compared to the state of the art works (MSS [1], MPSS [3], MBMSS [7]), this is indicated by the improved performance in terms of SNR.
2. FPGA based hardware design and implementation of the proposed speech enhancement methodology.

This paper is organized as follows. In section 2, a brief background of magnitude spectral subtraction and multi band spectral subtraction algorithms are presented; our proposed hardware design for speech enhancement architecture is presented in Section 3; in Section 4 hardware implementation and in Section 5 performance analysis are presented; concluding remarks in Section 6.

2 Background

In this section we discuss some of the fundamental issues related to spectral subtraction technique that are extremely important to understand our work presented in the next subsequent sections.

2.1 Spectral Subtraction Algorithm

Spectral subtraction is a procedure for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noisy signal in time domain is represented as:

$$y(m) = x(m) + n(m) \quad (1)$$

where $y(m)$, $x(m)$ and $n(m)$ are the signal, additive noise and the noisy signal respectively and m is the discrete time index.

The frequency domain noisy signal model corresponding to equation (1) can be represented as:

$$Y(f) = X(f) + N(f) \quad (2)$$

Where $Y(f)$, $X(f)$ and $N(f)$ are the frequency domain signals corresponding to $y(m)$, $x(m)$ and $n(m)$ respectively.

The noise estimation filter calculates $N(f)$ from the noisy spectrum. The magnitude of $N(f)$ is calculated by its average value during non speech activity. Spectral error [14] comes from subtraction estimator. It reduces by simple modification like magnitude averaging, half wave rectification, residual noise reduction and additional signal attenuation during non speech activity.

The discontinuities at the end point of the segment can be done by windowing of the signal and can be expressed as:

$$y_{\omega}(m) = x_{\omega}(m) + n_{\omega}(m). \quad (3)$$

Windowing signal can be expressed in frequency domain as:

$$Y_{\omega}(f) = W(f) * Y(f) = X_{\omega}(f) + N_{\omega}(f) \quad (4)$$

where the operator $*$ denotes convolution.

A scaled estimate of the magnitude spectra of the noise signal $\hat{N}_\omega(f)$ is subtracted from the corresponding spectra of the noisy signal $Y_\omega(f)$ to estimate the clean voice $\hat{S}_\omega(f)$,

$$|\hat{S}_\omega(f)|^\gamma = |Y_\omega(f)|^\gamma - |\alpha \hat{N}_\omega(f)|^\gamma \quad (5)$$

Noise signal is estimated and the frequency dependent subtraction factor α is included to compensate the overestimation of the instantaneous noise spectrum. $\gamma = 1$ for the magnitude spectral subtraction and $\gamma = 2$ for power spectral subtraction. The enhanced signal spectrum is obtained using the magnitude estimate $\hat{S}(f)$ and phase $\phi(f)$ of the corrupted input signal,

$$\hat{S}(f) = |\hat{s}f|e^{j\phi(f)} \quad (6)$$

Finally, the clean signal is obtained by the inverse fourier transform of $\hat{S}(f)$,

$$\hat{s}(m) = F^{-1}\hat{S}(f) \quad (7)$$

This general spectral subtraction method provides better results of the speech enhancement for high SNR signals compared to the low SNR signals. A combination of magnitude and phase spectral subtraction methods provides speech enhancement of both high and low SNR signals [3].

2.2 Multi Band Spectral Subtraction Algorithm

Most of the real world noise are colored noise, which affects the signal at various time interval. Multi band spectral subtraction algorithm [7] provides subtraction over individual frequency bands for the better speech enhancement.

The clean speech spectrum of equation (2) can be represented as:

$$|\hat{S}_\omega(f)|^\gamma = |Y_\omega(f)|^\gamma - \alpha |\hat{N}_\omega(f)|^\gamma \quad (8)$$

where α is the over subtraction factor, which is the function of the segmental SNR. General spectral subtraction methods assume that the noise is affected uniformly and the over subtraction factor α is subtracted over the whole spectrum. In real world the noise is effected in random phenomenon. The colored noise affects the signal spectrum differently at various frequencies. So the segmental SNR values change at different frequency bands [7]. The change of estimated SNR value for four frequency bands is shown in Fig. 1. This four frequency bands are linearly spaced.

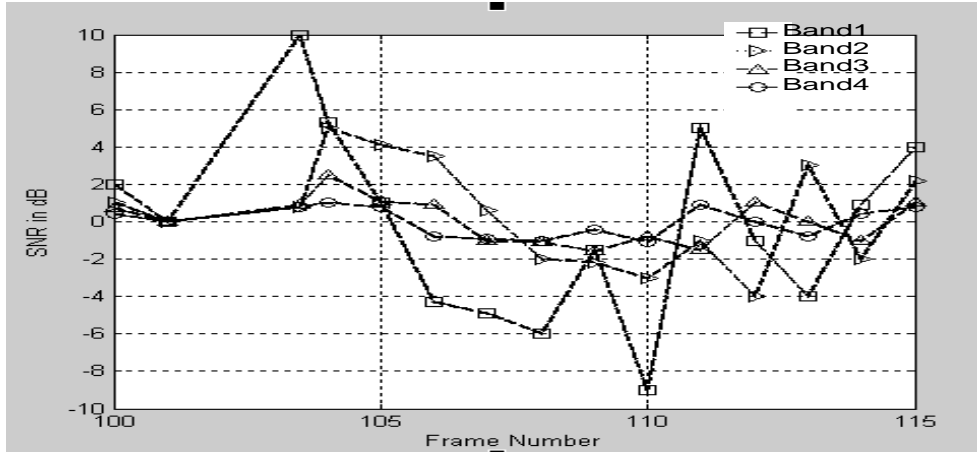


Figure 1: Segmental SNR comparison for different frequency bands

The speech spectrum is divided into different non overlapping bands and the subtraction procedure is done over each band independently. So, the enhanced signal spectrum of the i th frequency bands is,

$$|\hat{S}_i\omega(f)|^\gamma = |Y_i\omega(f)|^\gamma - \alpha_i \delta_i |\hat{N}_i\omega(f)|^\gamma \quad (9)$$

where α_i is the over subtraction factor of the i th frequency band and δ_i is the tweaking factor of each i th band. b_i and e_i are the beginning and ending frequency bins of i th frequency band. The over subtraction factor α is directly depended on the segmental SNR of the signal and is calculated as:

$$SNR_i(db) = 10 * \log_{10} \sum_{f=b_i}^{e_i} (|Y_i\omega(f)|/|N_i\omega(f)|)^2 \quad (10)$$

Depending upon the SNR_i value the over subtraction factor α_i evaluated as:

$$\alpha_i = \begin{cases} 5 & SNR_i < 5 \\ 4 - 3/20(SNR_i) & -5 < SNR_i < 5 \\ 1 & SNR_i > 20 \end{cases} \quad (11)$$

The over subtraction factor has a control in subtraction for each of the frequency bands. The subtraction factor δ_i provides an additional degree of control for each frequency bands. The value of δ_i [7] is specified by the following equation:

$$\delta_i = \begin{cases} 1 & f_i < 1KHz \\ 2.5 & 1KHz < f_i < FS/2 - 2KHz \\ 1.5 & f_i > FS/2 - 2KHz \end{cases} \quad (12)$$

Where f_i is the upper frequency band and FS is the sampling frequency.

After the subtraction of the estimated noise from all the frequency bands with different segmental SNR, we create the enhanced frequency bands.

3 Proposed Hardware Design

From the above discussion we observe that the MSS technique enhances the high SNR signals, MBMSS technique enhances the high SNR signals for different frequency bands. We propose a novel MPMBSS technique, which enhances both high SNR and low SNR signals for the different frequency bands.

In our proposed design we have four principle blocks namely magnitude multi band separation block, magnitude noise estimation-subtraction block, phase multi band separation block and phase noise estimation-subtraction block. The magnitude and phase operations are executed in parallel. The proposed architecture is shown in Fig. 2.

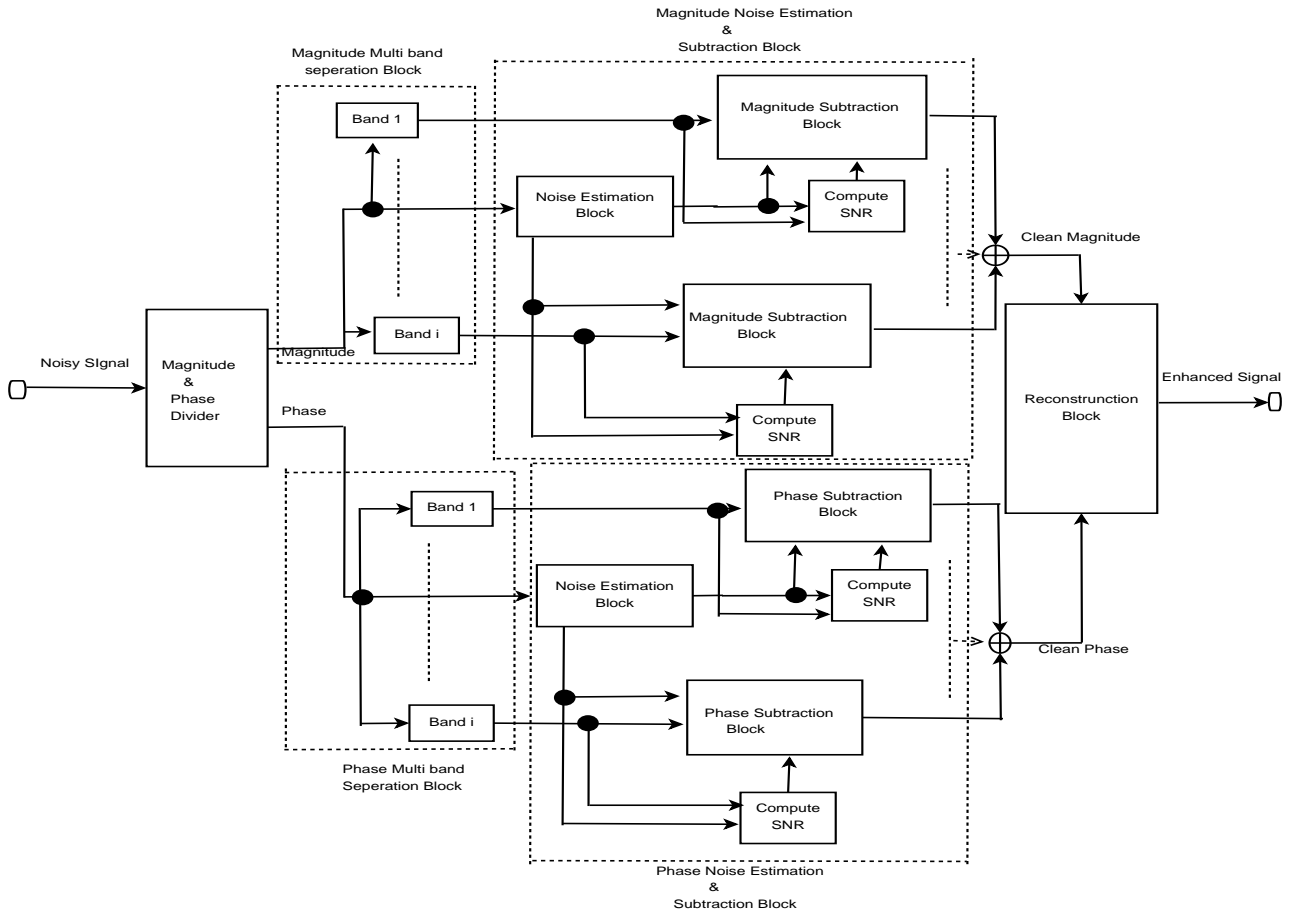


Figure 2: Proposed Architecture

From equation (9) we have,

$$|\hat{S}_i\omega(f)|^\gamma = |Y_i\omega(f)|^\gamma - \alpha_i\delta_i|\hat{N}_i\omega(f)|^\gamma \quad (13)$$

Where α_i is the over subtraction factor of the i^{th} frequency bands. The magnitude and the phase spectrum of the signal is divided into different frequency bands and the subtraction is done on each frequency bands from the estimated noise of both magnitude and phase spectrum of the signal. The noise estimated from the magnitude and noise spectrum of the signal are $\hat{N}_i\omega mg(f)$ and $\hat{N}_i\omega ph(f)$ respectively and it is subtracted from the noisy magnitude and phase spectrum with different frequency bands depending upon the α_i .

$$|\hat{S}_i\omega mg(f)|^\gamma = |Y_i\omega mg(f)|^\gamma - \alpha_i\delta_i|\hat{N}_i\omega mg(f)|^\gamma \quad (14)$$

$$|\hat{S}_i\omega ph(f)|^\gamma = |Y_i\omega ph(f)|^\gamma - \alpha_i\delta_i|\hat{N}_i\omega ph(f)|^\gamma \quad (15)$$

Where $|\hat{S}_i\omega mg(f)|$ and $\hat{S}_i\omega ph(f)$ are the clean magnitude and phase spectrum of the i^{th} frequency bands respectively. $|Y_i\omega mg(f)|$ and $Y_i\omega ph(f)$ are the noisy magnitude and phase spectrum of the i^{th} frequency bands respectively. And α_i is calculated from the compute SNR block.

The enhanced magnitude spectrum $|\hat{S}_i\omega mg(f)|$ and the enhanced phase spectrum $|\hat{S}_i\omega ph(f)|$ of the signal are combined together at the time of reconstruction of the signal.

$$\hat{S}_{mp}(f) = F^{-1}\hat{S}_{mp}(f) \quad (16)$$

$\hat{S}_{mp}(f)$ is the enhanced speech signal of the noisy signal.

We claim that the proposed design has an increased signal to noise ratio (SNR) as compared to the other existing state of the art architectures. The magnitude and phase operations are executed concurrently exploiting the parallel logic blocks of field programmable gate array (FPGA).

4 Hardware Implementation

The proposed architecture is implemented on the reconfigurable FPGA hardware. The time domain noisy speech signal is converted into the frequency domain signal by the Fast Fourier Transform (FFT) block. This signal is divided into magnitude and phase components using CORDIC ARCTAN DSP block. From the magnitude and phase spectrum, noise is estimated by the noise estimation block. The magnitude and phase spectrum are divided into four frequency bands [7] by multi band separation block and over subtraction factor (α_i) is calculated from each of the bands by using the SNR computation block. Magnitude and phase subtraction block subtracts the estimated noise from each of the frequency bands with different α_i . Enhanced magnitude and phase spectrum of the different frequency bands are combined together by the adder block and thus generating the enhanced magnitude and phase spectrum of the signal.

The enhanced phase signal passes through the CORDIC SINCOS to generate the real and imaginary form of the phase spectrum. These real and imaginary phases are combined with the enhanced magnitude spectrum using the multiplier block and is passed through the inverse FFT (IFFT) block to reconstruct the signal. Output of the IFFT block gets the enhanced speech signal. The block diagram of the proposed hardware design is shown in Fig. 3.

4.1 Fast Fourier Transform

The real world noisy speech signals are in the time domain signals. To convert this time domain signal to frequency domain we require the fourier transform over the signal. Fast Fourier Transform (FFT) block is used to perform the fourier operation on the signal. The FFT block generates the real as well as the imaginary components of the signal. We have used the FFT block of the Xilinx system generator platform. The option input/output was chosen for the FFT to implement its pipelined versions. For the performance optimization of the FFT block 4 multiplier structures are used and the phase factor is set to 8. Also the signal is segmented on non overlapping window of 256 samples. The data is recorded in the 2 stages of the block RAM (BRAM). The interface of the FFT block is shown in Fig. 4a. The real and imaginary components are transferred to the magnitude and phase separation block.

4.2 Magnitude and Phase Separation Block

Output of the FFT block drives the CORDIC ARCTAN block where the real and imaginary signal are divided into their magnitude and phase format. The magnitude $Y_{\omega mg}(f)$ and the phase $Y_{\omega ph}(f)$ are passed through the magnitude noise estimation block and phase noise estimation block respectively.

In the cordic ARCTAN block [25], architectural configuration is set to parallel mode for high throughput. The pipeline mode is set to maximum and the phase format is set to radians mode. The output width of this

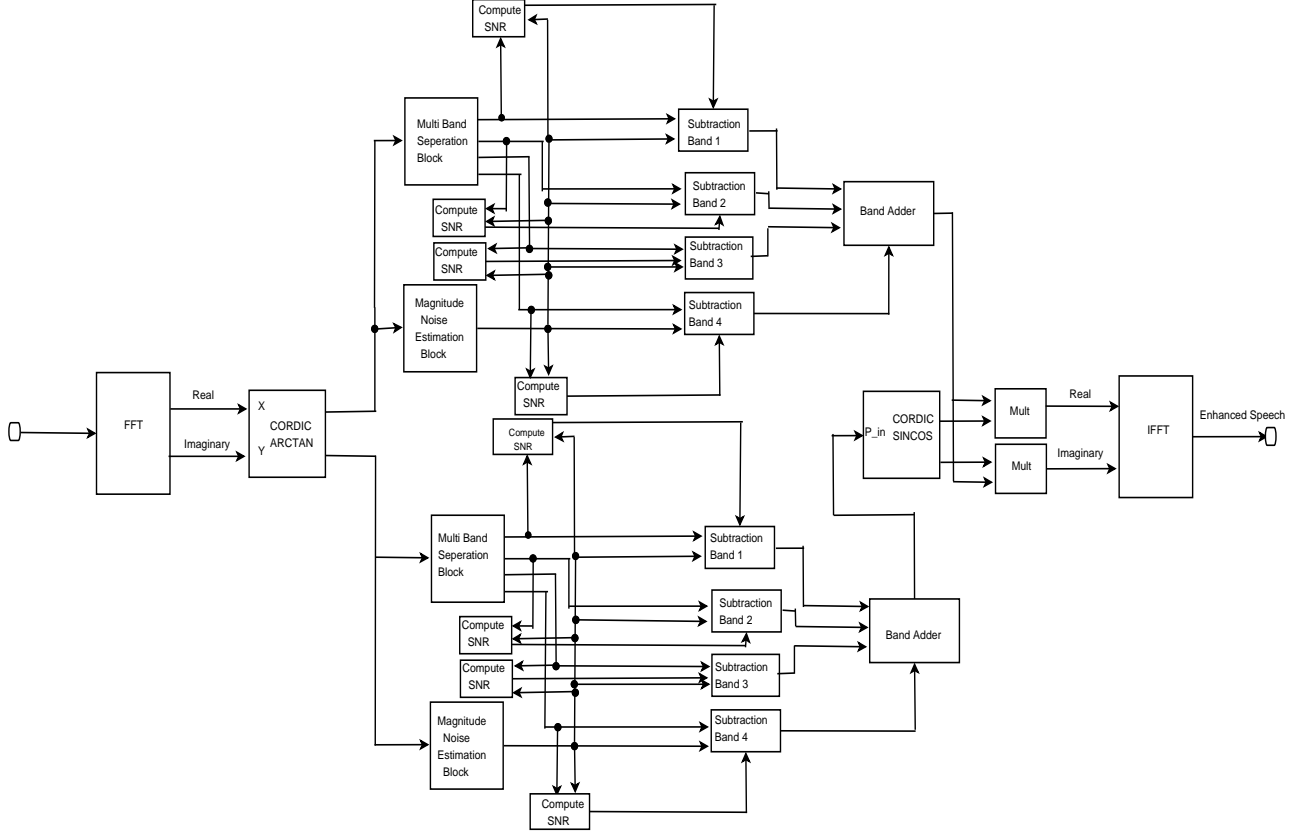


Figure 3: Hardware Design

block is set to 16. The output pins are magnitude, phase and ready pin of the Cordic ARCTAN block. The block diagram of the magnitude and phase separation block is shown in Fig. 4b.

The magnitude and phase spectrum are fed to the magnitude noise estimation block and phase noise estimation block respectively in parallel, where the band separation, noise estimation and subtraction processes are done.

4.3 Noise Estimation Block

In the noise estimation blocks, noise is estimated during the first few samples where noise is only present. Our design is adaptive in nature with the only constraint that a few initial samples of the input signal for a duration of $1.25ms$ is only noise, which is a fair constraint for speech communication. The block diagram of the magnitude and phase noise estimation block is shown in Fig. 5a.

Magnitude spectrum and phase spectrum of the signal are passed to the magnitude noise estimation block

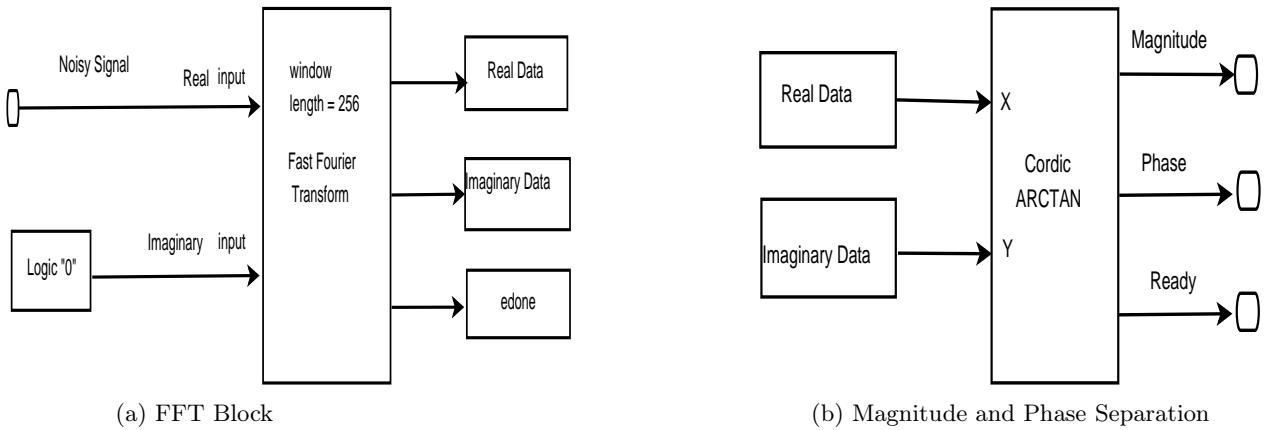


Figure 4: Block Diagram of FFT and CORDIC ARCTAN

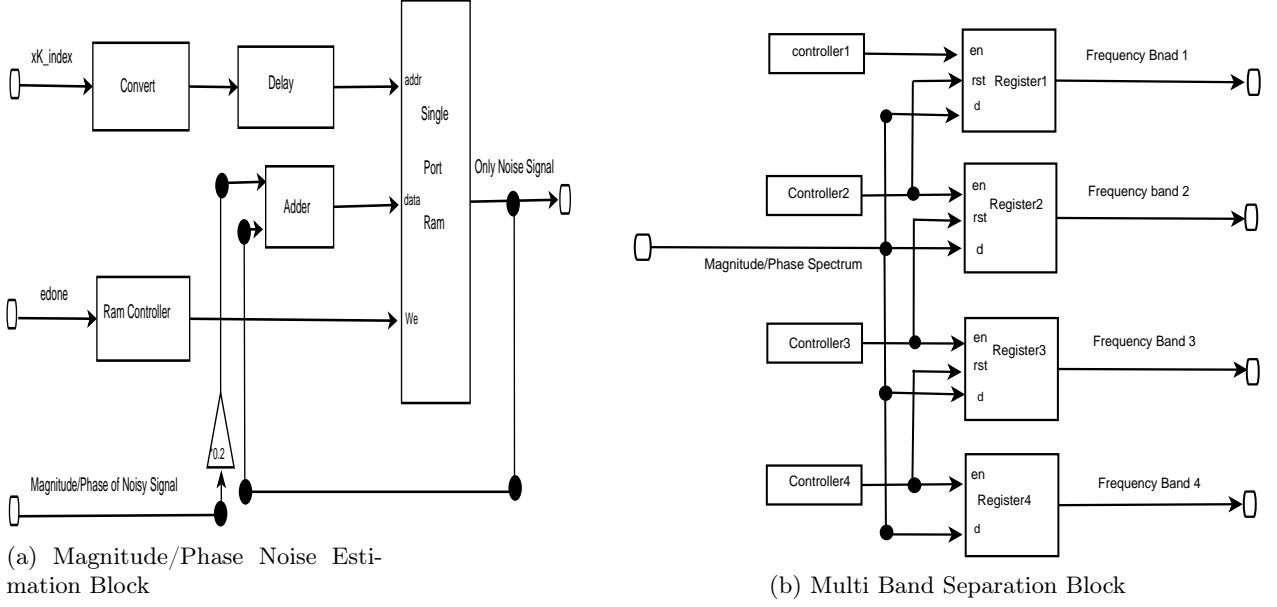


Figure 5: Block Diagram of Noise Estimation and Multi Band Separation Block

and phase noise estimation block respectively. In this design, one single port RAM is used to calculate noise power. RAM is set in write before read mode. Write enable pin of the single port RAM is active only for initial few samples (here 5 samples), which is taken care by the RAM controller block. One counter and one relational block are used to architect the RAM controller block. The first 5 samples of data are written on the RAM block and the other next samples are passed through the RAM block in the read mode. The noise samples that are stored in the RAM block are ready to be subtracted from the different frequency bands having magnitude spectrum and phase spectrum.

4.4 Multi Band Separation Block

The magnitude spectrum and phase spectrum of the signal also pass through multi band separation block. The magnitude spectrum and phase spectrum is divided into four frequency bands, which are linearly spaced. Four registers store the sample value of the four frequency bands until the reset signals are disabled. The multi band separation block is shown in Fig. 5b.

There are four controller blocks (Fig. 5b) that are used to divide the signal linearly into four frequency bands of the single signal. Each controller is connected to the enable pin of the registers. The reset pins of the first three register are enabled when the immediate next register enable pins are enabled. The fourth register have no reset pin, so the fourth register carry the rest of the signal. When the signal are in the first register by the controller1, the reset pin is disabled and the sample signal values are passed through the register and falls into the first subtraction block where the subtraction procedure is done. When controller2 becomes active then register2 is ready to accept signal and also register1 is in the reset mode by controller2 to avoid override the signals. When the register4 becomes active then rest of the signal passes through this register. If we require more frequency bands, then we need to connect more registers and controllers. But, we investigate in our design that four frequency bands of the signal give better throughput. This architecture is modeled in parallel configuration.

After subtraction of each bands from the estimated noise spectrum we reconstruct the signal into a single band frequency. Here we used multi band adder block to reconstruct the four signals. But before addition of the four enhanced signals, same controllers and registers which have been used for the separation are used for the same time format of the original signals.

4.5 Signal to Noise Ratio Computation

For the proposed architecture over subtraction factor α may vary for different frequency bands and depends on the signal to noise ratio for each frequency bands. The variation of α is described in equation (11). The architecture of calculating SNR is shown in Fig. 6a.

Maximum amplitude of signal and noise are calculated by the relational block, multiplexer and register block. The enable pin of the register is connected to the output of the controller to keep the same bandwidth of the divided signal. The controller blocks are the same which was used previously to separate frequency

bands. Those four controllers are used to the SNR computation block to get the maximum amplitude of the signal and noise. The output of the multiplexer holds the maximum value for each sample comparison. The registers pass the final maximum value of the signal samples by the controller block. Maximum amplitude of signal and noise pass through a division block to get the SNR value. In our proposed architecture, we require four SNR computation block for magnitude spectrum and four for the phase spectrum to estimate SNR values. The over subtraction factor α_i of each bands are calculated using the corresponding SNR values. Then subtraction done between each frequency bands and estimated noise with the different α_i .

4.6 Subtraction Block

In the subtraction block, estimated noise spectrum with over subtraction factor is subtracted from the signal spectrum of each of the frequency bands. The enhanced spectrum of i th frequency band is,

$$|\hat{S}_i\omega(f)|^\gamma = |Y_i\omega(f)|^\gamma - \alpha_i\delta_i|\hat{N}_i\omega(f)|^\gamma \quad (17)$$

The architecture of the proposed subtraction block is shown in Fig. 6b.

A multiplier block is used to combine the estimated noise spectrum and α which is depended on the signal to noise ratio. The subtraction block is used to get the enhanced signal spectrum. The final enhanced spectrum is in the output of the multiplexer. Our proposed architecture requires four subtraction block for magnitude subtraction and four for the phase subtraction as shown in Fig. 3.

4.7 Signal Reconstruction Block

Enhanced magnitude and phase spectrum are combined together to reconstruct the signal. Reconstruction of signal requires the inverse fourier transform over the magnitude and phase spectrum. The enhanced phase spectrum of the signal pass through the CORDIC SINCOS block [25] to get their real and imaginary format. Two multiplier blocks are used to multiply the enhanced magnitude spectrum of the real and imaginary enhanced phase spectrum. The IFFT block used to reconstruct the signal and provide the enhanced speech signal. The reconstruction process of the signal is demonstrate in Fig. 6c.

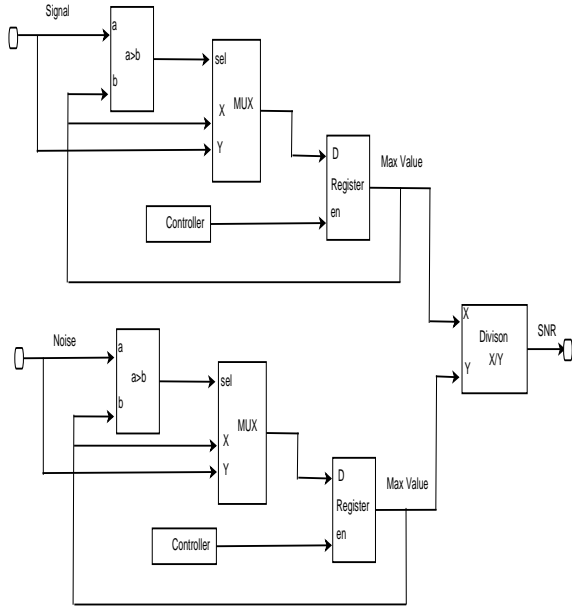
IFFT block is used to perform the inverse fourier transform operation over the signal. The option input/output was chosen for the IFFT to implement its pipelined versions. For the performance optimization of the FFT block, 4 multiplier structures are used and the phase factor is set to 8. Also the signal is segmented on non overlapping window of 256 samples. The architectural configuration of cordic SINCOS block is set to parallel mode for high throughput. The pipeline mode is set to maximum and the phase format is set to in the radians mode. The output width of this block is set to 16.

5 Performance Analysis

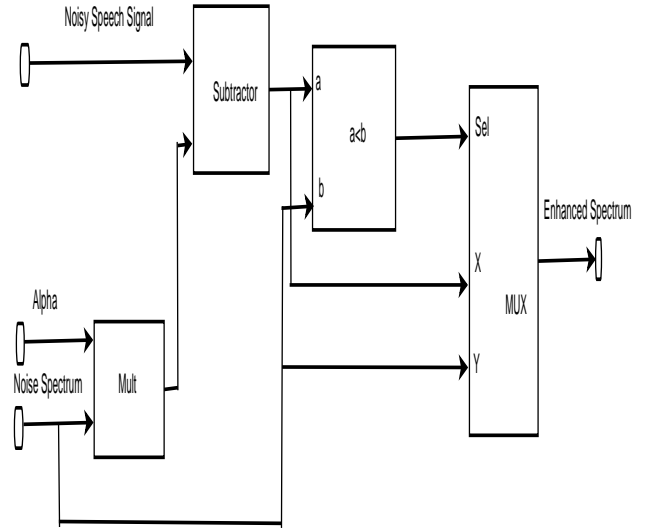
Field Programmable Gate Array (FPGA) contains a matrix of re-configurable logic circuitry. Different operations do not have to compete for the same processing resources because of the available special parallelism. So multiple control loops can run on a single FPGA device at different rates. The re-configurability of FPGAs can provide limitless flexibility. Most real-time systems require fast processing, which are met by the present day high speed FPGAs. The above mentioned hardware execution has been carried out on Atlys Spartan 6 FPGA board (Xilinx Spartan-6 LX45 FPGA, 324-pin BGA package, 128Mbyte DDR2 16-bit wide data). Spartan-6 LX FPGAs are optimized for applications that require the absolute lowest cost. It provides up to 150K logic cells, integrated PCI express blocks, advanced memory support, 390MHz DSP slices, and 3.2 Gbps low-power transceivers.

Here, we have used all the sound sources from [23] [24] except market noise, railway platform noise and train horn noise shown in Table II. This market, railway platform and train horn noises has been recorded from the respective environment. Football ground, market, car, railway platform, train horn and exhibition hall noisy signal was sampled at 16000 Hz and white, pink, cockpit, wind and factory noise are sampled at 8000 Hz. Due to the parallel nature of the proposed architecture and avoid sequential execution in software platform, we are implemented our design in FPGA. To compare our design with existing works [1] [7] [3], we have implemented [1] [7] [3] in FPGA because no such hardware implementation was found in the respective literature.

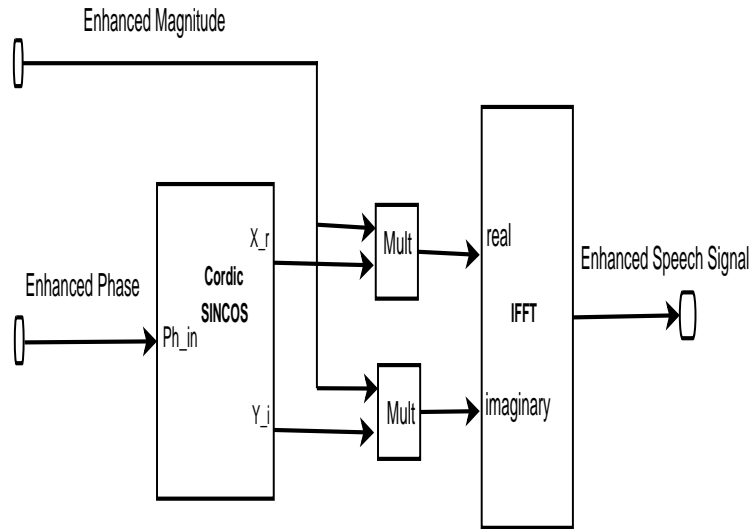
The device utilization of our implementation is shown in Table 1.. In Table 3, gives the system delay where only magnitude or phase operations are taking into account due to their parallel nature. Overall system unit delay of this design is 604 and execute in Xilinx Spartan-6 LX45 FPGA. The time requirement for execution of the proposed design on Xilinx Spartan-6 LX45 FPGA board is 6.04 *microseconds* where the board clock frequency is 100MHz. In Table 2, we compared the signal to noise ratio for all four methods from low



(a) Signal to Noise Ratio Computation

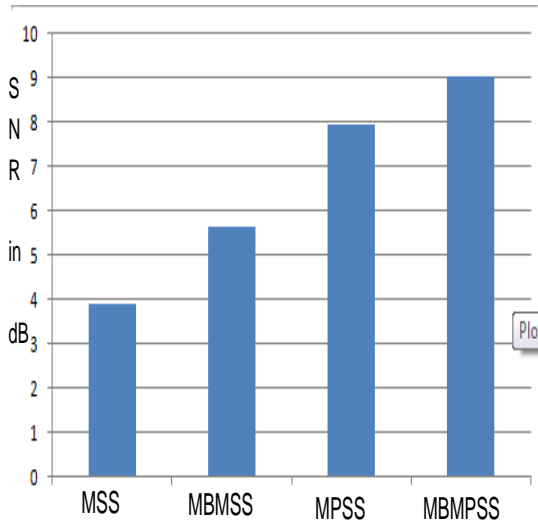


(b) Subtraction Block

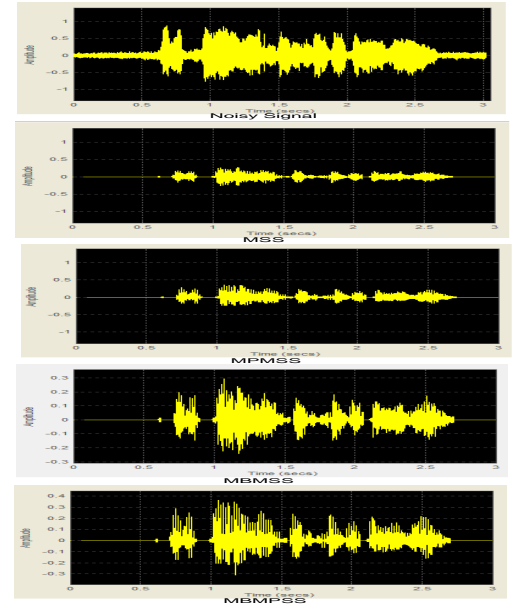


(c) Inverse Fourier Transform Block

Figure 6: Block Diagram of Signal to Noise Ratio Computation, Subtraction and Inverse Fourier Transform Block



(a) Rating of MSS, MBMSS, MPSS, MBMPSS



(b) Scope representation of MSS, MPSS, MBMSS, MBMPSS

Figure 7: Rating And Scope Representation of MSS, MPSS, MBMSS, MBMPSS

SNR to high SNR. Rating of MSS, MBMSS, MPSS and MBMPSS are shown in Fig. 7a. We observed that proposed method provides best result in every case in all SNR conditions. The melioration of the train horn noise was not outstripped due to high baseline. Time scope representation of the hardware implementation of MSS, MBMSS, MPSS and MBMPSS are shown in Fig. 7b. So, we can conclude that our proposed design significantly outstrip the other existing methods in terms of SNR mainly.

Table 1: Device utilization for SPARTAN 6 LX 45 FPGA

Device utilization summary	Available	used	utilization(%)
Slice Registers	184,304	8451	4
Slice LUTs	92,152	7544	8
Slice memory	21,680	1,161	5
Bonded IOBs	296	42	14
DSP48A1S	180	43	23

6 Conclusion

In this paper, we have proposed a novel hardware design for speech enhancement based on the spectral subtraction algorithm. The subtraction procedure is performed on both magnitude and phase spectrum of the different frequency bands. In this way we are able to eliminate noise from high SNR signals as well as low SNR signals for the different frequency bands. FPGA based hardware implementation of the proposed architecture gives better performance in-terms of SNR and throughput over the existing architectures of MSS, MPSS, MBMSS.

Acknowledgment

This work has been supported by the University Grant Commission (UGC) RGNF-2012-13-SC-WES-26014, Govt of India.

Table 2: SNR compression

Input	input SNR	MSS [1]	MBMSS [7]	MPSS [3]	MBMPSS(Proposed)
White Noise	-3	0.53	1.79	3.95	5.01
	0	3.93	5.66	7.97	9.03
	3	6.99	8.01	11.07	12.25
	8	11.02	12.95	15.25	16.83
	10	13.62	14.96	16.13	17.69
Pink Noise	-3	3.10	4.58	7.63	8.91
	0	2.95	4.16	7.21	8.98
	3	5.59	7.13	10.01	11.28
	8	10.56	12.01	14.51	15.93
	10	12.65	14.14	16.72	17.98
Cockpit Noise	-2	0.67	2.35	5.08	6.16
	0	2.54	4.08	6.75	7.73
	2	4.32	5.23	7.78	8.26
	6	8.92	9.76	12.12	13.22
	10	12.91	13.96	16.83	17.99
Football Ground Noise	-5	-1.11	0.02	2.88	4.01
	0	3.64	4.84	6.89	7.81
	5	8.95	10.09	12.27	13.12
	9	12.25	13.39	16.25	17.34
	13	16.42	17.58	20.39	21.48
Exhibition Hall Noise	-3	0.48	1.77	4.61	5.73
	0	3.34	4.47	7.38	8.30
	3	6.21	7.49	10.42	11.25
	8	11.39	12.55	15.34	16.44
	12	15.42	16.56	19.51	20.47
Market Noise	-3	0.42	1.60	4.44	5.56
	0	3.39	4.55	7.47	8.35
	3	6.41	7.57	10.49	11.38
	8	11.36	12.47	15.35	16.26
	12	15.56	16.79	19.78	20.89
Railway Platform Noise	-5	-1.08	0.15	3.01	3.92
	0	3.43	4.69	7.56	8.49
	5	8.35	9.47	12.38	13.26
	10	13.46	14.55	17.51	18.38
	13	16.39	17.52	20.30	21.11
Wind Noise	-3	0.39	1.26	4.53	5.47
	0	3.47	4.64	7.32	8.61
	3	6.15	7.56	10.36	11.34
	8	11.46	12.43	15.27	16.36
	13	16.37	17.43	20.28	21.07
Car Noise	-3	0.49	1.76	4.53	5.61
	0	3.48	4.65	7.66	8.46
	3	6.54	7.69	10.52	11.49
	8	11.44	12.63	15.57	16.48
	12	15.69	16.93	19.84	20.98
Factory Noise	-5	-1.36	0.42	3.58	4.32
	0	3.27	4.41	7.63	8.84
	5	8.52	9.73	12.59	13.61
	10	13.71	14.83	17.67	18.64
	13	16.56	17.68	20.46	21.24
Bursting Noise	-3	0.36	1.27	3.65	4.92
	0	3.38	5.46	7.62	8.89
	3	6.74	7.88	10.79	12.04
	8	10.93	12.85	15.14	16.69
	10	13.47	14.71	15.93	17.42
Train Horn Noise	-3	12.37	13.66	16.74	17.98
	0	15.16	16.76	19.38	20.51
	3	18.32	19.29	22.10	23.03
	8	22.96	23.77	25.86	26.64
	12	25.74	26.13	28.75	29.43

Table 3: System Delay

Hardware architecture	Delay
FFT	278
Magnitude-phase Extraction	13
Magnitude/phase Operation	24
Cordic SinCos	11
IFFT	278

References

- [1] Boll, S., "Suppression of acoustic noise in speech using spectral subtraction," in *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 1979, vol. 27., no. 2, pp. 113–120, doi = 10.1109/TASSP.1979.1163209., ISSN = 0096 – 3518.
- [2] Berouti, M. and Schwartz, R. and Makhoul, J., "Enhancement of speech corrupted by acoustic noise" *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.*, 1979, vol. 4., pp. 208–211, doi = 10.1109/ICASSP.1979.1170788.
- [3] Yi Zhang, Yunxin Zhao, "Real and imaginary modulation spectral subtraction for speech enhancement", *Speech Communication*, Volume 55, Issue 4, May 2013, Pages 509-522, ISSN 0167-6393, <http://dx.doi.org/10.1016/j.specom.2012.09.005>.
- [4] Lin, L., W. H. Holmes, and E. Ambikairajah. "Adaptive noise estimation algorithm for speech enhancement." *Electronics Letters* 39.9 (2003): 754-755.
- [5] Evans, N.W.D.; Mason, J.S.D.; Liu, W.M.; Fauve, B., "An Assessment on the Fundamental Limitations of Spectral Subtraction," *Acoustics, Speech and Signal Processing.*, 2006. IEEE International Conference on , vol.1, no., pp.I,I, 14-19 May 2006.
- [6] Lu, Y., Loizou, P., 2008. A geometric approach to spectral subtraction. *Speech Comm.* 50, 453–466.
- [7] Kamath, Sunil; Loizou, Philipos, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *Acoustics, Speech, and Signal Processing (ICASSP)*, 2002 IEEE International Conference on , vol.4, no., pp.IV-4164,IV-4164, 13-17 May 2002 doi: 10.1109/ICASSP.2002.5745591
- [8] Chao Li; Wen-Ju Liu, "A novel multi-band spectral subtraction method based on phase modification and magnitude compensation," *Acoustics, Speech and Signal Processing (ICASSP)*, 2011 IEEE International Conference on , vol., no., pp.4760,4763, 22-27 May 2011 doi: 10.1109/ICASSP.2011.5947419
- [9] Aarabi, P.; Guangji Shi, "Phase-based dual-microphone robust speech enhancement," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* , vol.34, no.4, pp.1763,1773, Aug. 2004 doi: 10.1109/TSMCB.2004.830345
- [10] Adiono, T. and Purwita, AA and Haryadi, R. and Mareta, R. and Priandana, E.R., "A hardware-software co-design for a real-time spectral subtraction based noise cancellation system," *Intelligent Signal Processing and Communications Systems (ISPACS)*, pp. 5-10, Nov 2013, doi=10.1109/ISPACS.2013.6704513
- [11] Hu, H.T.; Yu, C., "Adaptive noise spectral estimation for spectral subtraction speech enhancement," *Signal Processing, IET* , vol.1, no.3, pp.156,163, September 2007 doi: 10.1049/iet-spr:20070008
- [12] E. Verteletskaya and B. Simak, "Noise reduction based on the modified spectral subtraction method," *IAENG International journal of computer science*, Feb 2011.
- [13] Chen j.; Benesty, J.; Yiteng H.; Doclo, S., "New insights into the noise reduction Wiener filter," *Audio, Speech, and Language Processing, IEEE Transactions on* , vol.14, no.4, pp.1218,1234, July 2006.
- [14] Ephraim, Y. and Malah, D., "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," in *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 1984, pp. 1109-1121, doi= 10.1109/TASSP.1984.1164453, ISSN=0096-3518.
- [15] Guoshen Y. and Mallat, S. and Bacry, E., "Audio Denoising by Time-Frequency Block Thresholding," in *Signal Processing, IEEE Transactions on*, 2008, vol. 56, no. 5, pp. 1830-1839, doi=10.1109/TSP.2007.912893, ISSN=1053-587X.

- [16] Yasser G. and Mohammad R. Karami-M., "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets," in *Speech Communication*, 2006, vol. 48, no. 8, pp. 927 – 940, doi=<http://dx.doi.org/10.1016/j.specom.2005.12.002>, url = <http://www.sciencedirect.com/science/article/pii/S0167639305002888>.
- [17] Jwu-Sheng Hu, Ming-Tang Lee, Chia-Hsing Yang, An embedded audio–visual tracking and speech purification system on a dual-core processor platform, *Microprocessors and Microsystems*, Volume 34, Issues 7–8, November 2010, Pages 274-284, ISSN 0141-9331, <http://dx.doi.org/10.1016/j.micpro.2010.05.004>.
- [18] Ka Fai Cedric Yiu, Zhibao Li, Siow Yong Low, Sven Nordholm, FPGA multi-filter system for speech enhancement via multi-criteria optimization, *Applied Soft Computing*, Volume 21, August 2014, Pages 533-541, ISSN 1568-4946, <http://dx.doi.org/10.1016/j.asoc.2014.03.016>.
- [19] Halupka, D.; Rabi, A.S.; Aarabi, P.; Sheikholeslami, A., "Low-Power Dual-Microphone Speech Enhancement Using Field Programmable Gate Arrays," *Signal Processing, IEEE Transactions on*, vol.55, no.7, pp.3526,3535, July 2007 doi: 10.1109/TSP.2007.893918
- [20] Chang-Min Kim; Hyung-Min Park; Taesu Kim; Yoon-Kyung Choi; Soo-Young Lee, "FPGA implementation of ICA algorithm for blind signal separation and adaptive noise canceling," *Neural Networks, IEEE Transactions on*, vol.14, no.5, pp.1038,1046, Sept. 2003 doi: 10.1109/TNN.2003.818381
- [21] McAllister, John, "FPGA-based DSP", *Springer US*, doi = 10.1007/978 – 1 – 4419 – 6345 – 1 – 14, pp. 363-392.
- [22] www.mathworks.com/products/hdl-verifier
- [23] <http://ecs.utdallas.edu/loizou/speech/noizeus>
- [24] <http://www.cslu.ogi.edu/nsl/data>
- [25] www.xilinx.com/support/sw-manual

